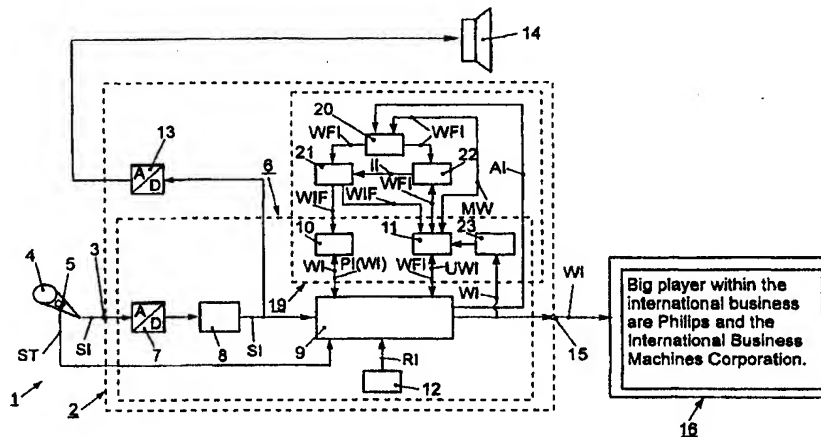




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : G10L 15/18		A1	(11) International Publication Number: WO 00/31727
			(43) International Publication Date: 2 June 2000 (02.06.00)
(21) International Application Number: PCT/EP99/08640 (22) International Filing Date: 11 November 1999 (11.11.99) (30) Priority Data: 98890348.0 24 November 1998 (24.11.98) EP (71) Applicant: KONINKLIJKE PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL). (72) Inventors: BARTOSIK, Heinrich; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). KLAKEOW, Dietrich, G.; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). (74) Agent: RÖGGLA, Harald; Internationaal Octrooibureau B.V., Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).		(81) Designated States: JP, KR, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published With international search report.	

(54) Title: OPTIMIZATION DEVICE FOR OPTIMIZING A VOCABULARY OF A SPEECH RECOGNITION DEVICE



(57) Abstract

In an optimization device (19) for optimizing a vocabulary of a speech recognition device (2), comprising a lexicon memory (10) in which word information (WI) of at least a first and a second word forming the vocabulary of a speech recognition device (2) can be stored, and comprising a speech model memory (11) in which at least a probability of occurrence of the second word after the first word can be stored as transition probability information (UWI) in a word sequence formed by these words, and comprising word defining means (21) for defining a third word and for storing the third word as word information (WI) in the lexicon memory (10) and for storing at least transition probability information (UWI) of the probability of occurrence of the third word in a word sequence after at least the first or the second word stored in the lexicon memory (10) in the speech model memory (11), test means (20) are provided which are arranged for testing whether transition probability information (UWI) of a word sequence stored in the speech model memory (11) has a minimum value (MW) and the word defining means (21) are arranged for defining the words of this word sequence as the third word when the test of the test means (20) shows a positive result.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Optimization device for optimizing a vocabulary of a speech recognition device.

The invention relates to an optimization device for optimizing a vocabulary of a speech recognition device, comprising a lexicon memory in which the word information of at least a first and a second word forming the vocabulary of a speech recognition device can be stored, and comprising a speech model memory in which at least a probability of occurrence
5 of the second word after the first word in a word sequence formed by these words can be stored as transition probability information, and comprising word defining means for defining a third word and for storing in the lexicon memory the third word as word information and for storing in the speech model memory at least transition probability information of the probability of occurrence of the third word in a word sequence after at least the first or the
10 second word stored in the lexicon memory.

The invention further relates to a speech recognition device for recognizing phoneme information contained in speech information of a spoken text and for delivering word information of a recognized text, comprising input means which are supplied with speech information of a spoken text as input signals, and comprising speech recognition means
15 which are arranged for recognizing phoneme information of the spoken text contained in the input signals and for delivering word information of a recognized text, and comprising output means which can deliver word information of a recognized text as output signals.

The invention further relates to a vocabulary generator for generating and storing word information that forms the vocabulary of a speech recognition device, comprising
20 input means to which word information of stored text information can be fed as input signals, and comprising generator means which are arranged for generating a vocabulary, so that at least word information of the text information of a first and a second word and transition probability information indicating the probability of occurrence of the second word after the first word in a word sequence of the text information can be determined, and which generator
25 means are arranged for storing at least the first and the second word as word information in a lexicon memory and the transition probability information in a speech model memory.

Such a speech recognition device for recognizing phoneme information contained in speech information of a spoken text and for delivering word information of a

recognized text, of the type discussed in the second paragraph, comprising an optimization device for optimizing a vocabulary of a speech recognition device of the type defined in the first paragraph, is known from document WO 96/29695.

The known speech recognition device includes input means formed by a
5 microphone terminal to which a microphone can be connected. The microphone can apply speech information as an electric input signal to the speech recognition device, which speech information comes from a text spoken by a user of the speech recognition device. The electric input signal can be applied to speech recognition means of the speech recognition device and the speech recognition means can deliver its recognized word information as recognized text
10 to output means of the speech recognition device. A monitor by which word information of the recognized text can be displayed can be connected to the output means, which are formed by a monitor terminal.

For recognizing word information contained in the electric input signal, the speech recognition means include, inter alia, a lexicon memory. The lexicon memory stores as
15 word information all the words that form the vocabulary of the speech recognition device. Phoneme information is assignedly stored with word information, which phoneme information forms a phoneme sequence featuring the assignedly stored word.

During a speech recognition operation of the speech recognition device, phoneme sequences in a spoken text are determined by the speech recognition means and
20 compared with phoneme sequences stored in the lexicon memory. If this comparison shows a match of a determined phoneme sequence and a stored phoneme sequence, stored word information assigned to this stored phoneme sequence is taken from the lexicon memory as a recognized word.

The speech recognition means further include a speech model memory which
25 stores transition probability information for word sequences of words stored in the lexicon memory. The speech model memory stores word sequences of two words each, so-called bigrams, and word sequences of three words each, so-called trigrams.

For example, the word sequence "Sehr geehrte Damen und Herren" made up of bigrams and trigrams relatively often occurs as a typical formulation in spoken texts. In the
30 speech model memory is stored as a probability of occurrence of the word "Damen" after the words "Sehr geehrte" the transition probability information of, for example, "5%". For the probability of occurrence of the word "Herren" after the words "Sehr geehrte" is stored the transition probability information of, for example, "4%". Since the composite word sequence "Sehr geehrte Damen und Herren" as a typical formulation occurs more frequently than the

composite word sequence "Sehr geehrte Herren" in a spoken text, the probability of occurrence of the trigram "Sehr geehrte Damen" is about "1%" higher than the probability of occurrence of the trigram "Sehr geehrte Herren".

During a speech recognition operation of the speech recognition device, a
5 search is made in the speech model memory not only for words separately stored in the lexicon memory, but also for word sequences formed by stringed recognized words. By evaluating transition probability information stored in the speech model memory for recognized word sequences, the word information of the composite word sequence is determined as recognized text by the speech recognition device, of which word information
10 the transition probability information of the bigrams and trigrams contained in the word information has the highest value.

It is known to the expert that typical formulations in spoken texts are better recognizable the larger the number of words are which are stored per word sequence in a speech model memory. For example, the composite word sequence "Sehr geehrte Damen und
15 Herren" could be recognized very well by a speech recognition device, if it was stored in the speech model memory as a word sequence having only one transition probability information signal with a very high value, because this word sequence would not have to be assembled from bigrams und trigrams by the speech recognition means during a speech recognition operation. Consequently, also the computation circuitry of the speech recognition means
20 would be relatively small, which would also be an advantage.

In this case, where word sequences stored in the speech model memory would consist of up to five words, a very large number of possible and sensible combinations of five words would be stored in the lexicon memory and, therefore, a multiplicity of word sequences would have to be stored in the speech model memory. The required memory space in the
25 speech model memory would therefore be very large and the speech recognition device would be expensive, which is a considerable disadvantage.

The number of different words forming the vocabulary of the speech recognition device and which can be recognized by the speech recognition device is again restricted as a consequence of the memory capacity of the lexicon memory. The known speech
30 recognition device includes word defining means by which so-called composite elements can be defined as words. Composite elements are words which relatively often occur as composite words, so-called composites, in a spoken text. The word defining means are arranged for storing transition probability information in the speech model memory which information

indicates the probability of occurrence of a composite element defined as one word in a word sequence after at least one further word stored in the lexicon memory.

Storing composite elements in the lexicon memory provides the advantage that not all possible composites formed by combinations of such composite elements need to be stored in the lexicon memory, so that the required memory space of the lexicon memory is reduced considerably.

The known speech recognition device now proves to have a disadvantage that typical formulations contained in a spoken text are not sufficiently well recognized during a speech recognition operation, because only transition probability information of word sequences having a maximum of three words can be stored in the speech model memory and this number of words per word sequence cannot, in essence, be increased as a result of the constriction of the limited memory capacity of the speech model memory. Furthermore, in the known speech recognition device, the recognition of typical formulations contained in a spoken text is additionally degraded by the inclusion of words forming composite elements, because a composite formed, for example, by three composite elements already forms a complete word sequence and no further information about words and word sequences neighboring this word sequence can be determined by evaluating transition probability information of this word sequence.

The disadvantages stated above have also turned up in a speech recognition device of the type defined in the second paragraph when a vocabulary stored as word information was used in a speech model memory, which vocabulary was generated by a vocabulary generator as mentioned in the third paragraph and stored in the speech model memory.

It is an object of the invention to eliminate the problems stated above and provide an improved optimization device of the type mentioned in the opening paragraph. This object is achieved with an optimization device of the type mentioned in the opening paragraph in that test means are provided which are arranged for testing whether transition probability information of a word sequence stored in the speech model memory has a minimum value, and in that the word defining means are arranged for defining the words of this word sequence as a third word when the test means give a positive test result.

In consequence, word sequences containing two or more words, which relatively often appear as a typical formulation in spoken texts or stored text information, can be stored in the lexicon memory as one word. Thus, transition probability information on word

sequences can be stored in the speech model memory, which word sequences contain words which are themselves determined by two or more words stored as one word in the lexicon memory. This brings in the advantage that the number of words per word sequence stored in the speech model memory need not be increased and, nevertheless, formulations with a larger
5 number of words of a word sequence can be recognized better and the speech model has a so to speak larger range in a word sequence formed by recognized words.

Additionally, there is the further advantage that when a speech recognition is in operation, the speech recognition means are capable of determining phoneme sequences contained in a spoken text more reliably. Each phoneme in a phoneme sequence is influenced
10 in the way it is pronounced by phonemes before and after the particular phoneme in the phoneme sequence. The phonemes of a phoneme sequence of a word are influenced on the word boundaries by adjacent words in the word sequence. Consequently, in case of a recognition of a word sequence comprising a larger number of words, a larger number of neighboring phonemes of neighboring words is known and, as a result, the associated word
15 sequence can be recognized more reliably.

In an optimization device as claimed in claim 1 it has appeared to be advantageous to provide the measures in accordance with claim 2. Consequently, the advantage is obtained that transition probability information already determined of word sequences stored in the speech model memory, which contain one word defined by the word
20 defining means plus two or more words, can be stored as transition probability information for word sequences containing the defined word even after the word has been defined.

It is a further object of the invention to eliminate the problems stated above and provide an improved speech recognition device of the type as stated at the beginning of the application in the second paragraph. This object is achieved with a speech recognition device
25 of the type stated in the second paragraph in that an optimization device in accordance with claim 1 is provided.

As a result, the advantage is obtained that typical formulations in spoken texts can be recognized better and the recognition rate of the speech recognition device is improved, although practically no additional storage space is necessary for the speech model memory.
30 Furthermore, the computation circuitry of the speech recognition means during a speech recognition operation is considerably smaller when typical formulations are recognized, which is highly advantageous.

In a speech recognition device as claimed in claim 3 it has proved to be advantageous to provide the measures in accordance with claim 4. As a result, the advantage is

obtained that a word sequence defined as one word by the word defining means can be extended by further words contained in recognized texts before or after the defined word, so that the so-called range of the speech model is further extended.

5 It is a further object of the invention to eliminate the problems discussed above to provide an improved vocabulary generator of the type stated in the third paragraph. This object is achieved with a vocabulary generator of the type discussed in the third paragraph, in that an optimization device in accordance with claim 1 is provided.

As a result, the advantage is obtained that a vocabulary generated in the vocabulary generator and stored in the lexicon memory as word information is optimized to
10 the effect that the previously stated advantages are obtained with a speech recognition device.

In a vocabulary generator in accordance with claim 5 it has proved to be advantageous to provide the measures in accordance with claim 6. As a result, the advantage is obtained that a word sequence which the word defining means defined as one word can be extended by further words often contained in the stored text information before or after the
15 defined word, so that the so-called range of the speech model is further extended.

In a vocabulary generator as claimed in claim 5, it has proved to be advantageous to provide the measures in accordance with claim 7. As a result, the advantage is obtained that only those word sequences are defined as one word that occur relatively often in stored texts and thus also relatively often in spoken texts, so that the required memory space in
20 the speech model memory is further reduced.

These and other aspects of the invention are apparent from and will be elucidated with reference to the embodiments described hereinafter.

In the drawings:

25 Fig. 1 diagrammatically shows in the form of a block diagram a speech recognition device including test means for testing whether transition probability information stored in a speech model memory has a minimum value,

Fig. 2 shows a first table containing word information and phoneme information stored in a lexicon memory of the speech recognition device in accordance with
30 Fig. 1,

Fig. 3 shows a second table containing word sequence information and transition probability information stored in the speech model memory of the speech recognition device as shown in Fig. 1,

Fig. 4 shows a third table containing word information and phoneme information stored in the lexicon memory of the speech recognition device as shown in Fig. 1 which, after a word sequence comprising two words has been defined, is stored as one word by word defining means of the speech recognition device,

Fig. 5 shows a fourth table containing word sequence information and transition probability information stored in the speech model memory of the speech recognition device as shown in Fig. 1, which word sequence information and transition probability information is stored as one word after the word defining means have defined a word sequence containing two words,

Fig. 6 diagrammatically shows in the form of a block diagram a vocabulary generator including test means for testing whether transition probability information stored in a speech model memory has a minimum value.

Fig. 1 shows in the form of a block diagram a personal computer 1 in which a speech recognition device 2 in accordance with a first example of embodiment of the invention is realized. The speech recognition device is arranged for recognizing phoneme information PI contained in speech information SI of a text spoken by a user of the speech recognition device 2, and for delivering word information WI of a recognized text. The speech recognition device 2 includes input means formed by an input terminal 3.

A microphone 4 can be connected to the input terminal 3. The microphone 4 can deliver speech information SI of a spoken text as an electric input signal to the input terminal 3 of the speech recognition device 2. The microphone 4 has a control key 5 by which control information ST can be delivered to the speech recognition device 2.

When a user of the speech recognition device 2 wishes to speak a spoken text to be recognized into the microphone 4, the user is to actuate the control key 5. Subsequently, speech information SI contained in the spoken text can be delivered to the input terminal 3 and the control information ST to the speech recognition device 2.

The speech recognition device 2 includes speech recognition means 6 which are arranged for recognizing phoneme information PI of a spoken text, which phoneme information is contained in the speech information SI of the input signal, and for delivering word information WI of a recognized text. For this purpose, the speech recognition means 6 include an A/D converter stage 7, a storage stage 8, calculation means 9, a lexicon memory 10, a speech model memory 11 and a reference memory 12.

Speech information SI delivered to the input terminal 3 can be delivered to the A/D converter stage 7 as an electric input signal. The A/D converter stage 7 can deliver digitized speech information SI to the storage stage 8. The storage stage 8 stores the digitized speech information SI delivered thereto.

5 In an audio reproduction mode of the speech recognition device 2, which mode can be activated in a manner not further shown in Fig. 1, digitized speech information SI stored in the storage stage 8 can be applied to a D/A converter stage 13. The D/A converter stage 13, in the audio reproduction mode, can deliver analog speech information SI as electric output signals to a loudspeaker 14 for the acoustic reproduction of a text spoken into the
10 microphone 4 by a user of the speech recognition device 2.

The calculation means 9 are formed by a microprocessor and connected via an address/data bus to a lexicon memory 10, the speech model memories 11 and the reference memory 12. To the calculation means 9 can be applied digital speech information SI stored in the storage stage 8 and the control information ST coming from the microphone 4.

15 The calculation means 9 can determine word information WI of a recognized text while it utilizes information stored in the lexicon memory 10, the speech model memory 11 and the reference memory 12, which will be discussed in further detail hereinafter. The calculation means 9 can deliver word information WI of a recognized text to an output terminal 15 which forms output means. A monitor 16 on which word information WI of a
20 recognized text and delivered by the output terminal 15 can be displayed can be connected to the output terminal 15.

In the lexicon memory 10 can be stored the word information WI with a maximum of 64,000 individual words which form the vocabulary of the speech recognition device 2. The speech recognition device 2 correctly recognizes only the words contained in the
25 speech information SI of a spoken text that are also stored in the lexicon memory 10.

For each word information WI of a word in the lexicon memory 10 can be stored a phoneme sequence as phoneme information PI(WI) featuring the word. Phonemes of a phoneme sequence are the smallest distinguishable acoustic units into which digitized speech information SI can be subdivided. The acoustic pronunciation of a phoneme in a phoneme
30 sequence is influenced by the phonemes surrounding the relevant phoneme in the phoneme sequence. The first phoneme of a phoneme sequence of a word thus depends on the last phoneme of a phoneme sequence of a previous word in a word sequence, as also the last phoneme of a phoneme sequence of a word depends on the first phoneme of a phoneme sequence of the next word in the word sequence. In a speech recognition operation of the

speech recognition device 2 it is therefore very important for the correct recognition of a word to know the words surrounding the word of the word sequence to be recognized, or to adopt these words as predetermined values.

A first table 17 of Fig. 2 contains word information WI and phoneme information PI(WI) assignedly stored in the lexicon memory 10. For a simple explanation, the letters A, B, C to F are stated in the table 17 to represent the word information WI. The first table 17 for example contains for a word "international" the word information WI is A, for a word "machines" the word information WI = B, for a word "business" the word information WI = C, for a word "connection" the word information WI = D, for a word "the" the word information WI = E and for a word "corporation" the word information WI = F in brackets. The seven word information signals WI stated in the first table 17 represent a plurality of word information signals WI stored in the lexicon memory 10. The vocabulary of the speech recognition device 2 thus also contains the seven words denoted as word information WI in the first table.

A probability of occurrence of a second word stored in the lexicon memory 10 can be stored as transition probability information UWI in the speech model memory 11 of the speech recognition device 2, after a first word of a word sequence stored in the lexicon memory 11 in a word sequence formed by these words. In the speech model memory 11 can be stored word sequences of two words each, so-called bigrams and word sequences of three words each, so-called trigrams.

Fig. 3 shows a second table 18 containing word sequence information WFI of word sequences and assigned transition probability information UWI stored in the speech model memory 11. For example, the third row of the second table 18 contains the information that in a word sequence formed by the words "business" and "international" in speech information SI of a spoken text the word "business" having the word information WI = C follows the word "international" having the word information WI = A with a statistical probability of 10%. When during a speech recognition operation the word "international" is recognized, it may be assumed with a probability of 10% that the next word present in the spoken text will be the word "business", which in the previously discussed context, with the acoustic pronunciation of the last phoneme of the word "international" and of the first phoneme of the word "business", is very important for a correct recognition of the word.

Transition probability information UWI of word sequence information WFI = A+C+D of 2%, stated in the sixth row of the table 18, indicates that the word "connection" follows the word sequence "international business" with a probability of 2%. Transition

probability information UWI stated on the seventh row of the second table 18 of word sequence information WFI = E+A+C of 5% indicates that the word "business" follows the word sequence "the international" with a probability of 5%.

There may be observed that in the speech model memory 11 not the word
5 information WI such as in the lexicon memory 10 is again stored in the word sequence information WFI, but, to save storage space in the speech model memory 11, address pointers are stored at memory locations in the lexicon memory 10 of the relevant word information WI in the speech model memory 11. For example, an address pointer to the second row of the first table and an address pointer to the fourth row of the first table 17 are stored on the second row
10 of the third table 18 for the word sequence information WFI = A+C.

Reference information RI is stored in the reference memory 12. Since each human being has a different type of acoustic pronunciation of a word, also phonemes and phoneme sequences are pronounced slightly differently by each human being. The speech recognition device 2 is adapted to the respective user of the speech recognition device 2 by
15 means of reference information RI stored in the reference memory 12.

When a user speaks a text into the microphone 4, and simultaneously presses the control key 5, the control information ST delivered by the microphone 4 to the calculation means 9 activates a speech recognition mode in the speech recognition device 2 and a speech recognition operation in the calculation means 9. Speech information SI of the spoken text is
20 applied by the microphone 4 to the A/D converter stage 7 and from there as digitized speech information SI to the storage stage 8 and stored there.

When the calculation means 9 are ready for processing digitized speech information SI stored in the storage stage 8, the digitized speech information SI is read from the storage stage 8 by the calculation means 9. The calculation means 9, while utilizing the
25 reference information RI stored in the reference memory 12, determine phoneme information PI of phoneme sequences contained in the digitized speech information SI. The calculation means 9 then compare determined phoneme information PI with phoneme information PI(WI) stored in the lexicon memory 10. When there is a match between determined and stored phoneme information PI(WI) after this comparison, stored word information WI assigned to
30 this stored phoneme information PI(WI) is determined as a recognized word from the lexicon memory 10.

During the speech recognition operation of the calculation means 9, word sequences are formed by stringing together recognized words. The calculation means 9 compare word sequence information WFI of formed word sequences with word sequence

information WFI stored in the speech model memory 11. When there is a match, the stored transition probability information UWI is determined which is assigned to this recognized word sequence in the speech model memory 11.

By evaluating transition probability information UWI stored in the speech
5 model memory 11 for recognized word sequences, a plurality of possible word sequences composed of words and their overall transition probability information are determined. The word information WI of the composed word sequence is determined as recognized text, whose overall transition probability information of the bigrams and trigrams included therein, has the highest value. Such a speech recognition operation is carried out in accordance with the so-called "Hidden-Markov-Model" and has been known for a long time. Through the output
10 terminal 15 the calculation means 9 deliver word information WI of a recognized text recognized by the calculation means 9, to the monitor 16 for the recognized text to be displayed.

The speech recognition device 2 comprises an optimization device 19 which is
15 arranged for optimizing the vocabulary of the speech recognition device 2, which vocabulary is stored in the lexicon memory 10. For this purpose, the optimization device 9 includes test means 20, word defining means 21 and determining means 22. The calculation means 9 are arranged for activating an optimization mode of the speech recognition device 2 and an optimization operation of the optimization device 19 in that the calculation means deliver
20 activation information AI to the test means 20. The calculation means 9 could be arranged, for example, for activating an optimization operation of the optimization device 19 after a certain number of speech recognition operations or after, for example, one week since the last optimization operation. During an optimization operation of the optimization device 19, word sequences often occurring in spoken texts and stored as word sequence information WFI
25 stored in the speech model memory 11 are defined as one word to provide a better recognition of these typical formulations, which will be discussed in more detail hereinafter.

In the test means 20 are stored as a minimum value MW a value of 9% which prescribes that word sequences containing typical formulations, in which a second word occurs with an occurrence probability of at least 9% after a first word, are defined as one
30 word. By defining the minimum value MW, the required memory capacity to be expected for the speech model memory 11 and the computational circuitry in the calculation means 9 necessary for a typical formulation during a speech recognition operation can be predefined.

When activation information AI occurs, the test means 20 are arranged for comparing transition probability information UWI stored in the speech model memory 11 with

the minimum value MW. When transition probability information UWI stored in the speech model memory 11 has the minimum value MW, or a higher value, stored word sequence information WFI assigned to this transition probability information UWI in the speech model memory 11 can be determined from the speech model memory 11 by the test means 20. Such word sequence information WFI can be delivered by the test means 20 both to the word defining means 21 and to the determining means 22.

The word defining means 21 are arranged for defining as one word word sequence information WFI delivered thereto of a word sequence containing at least two words, and for storing word information WFI of a word sequence defined as one word as word information WI in the lexicon memory 10. As a result, the advantage is obtained that typical formulations in spoken texts can already be recognized as one word and not as a word sequence by means of information stored in the speech model memory 11. Advantageously, this considerably reduces the calculation circuitry of the calculation means 9 when a word sequence defined as one word is recognized.

In the case of a positive result of the test of the test means 20, the determining means 22 are arranged for comparing word sequence information WFI delivered to the determining means 22 by the test means 20 with the word sequence information stored in the speech model memory 11. If word sequence information WFI delivered to the determining means 22 corresponds to word sequence information WFI stored in the speech model memory 11, or is contained in stored word sequence information WFI, the determining means 22 are arranged for delivering identification information II to the word defining means 21.

Identification information II features a first memory location in the speech model memory 11 in which this word sequence information WFI is stored. However, identification information II also features all the further memory locations in the speech model memory 11 in which this word sequence information WFI contained in other word sequence information WFI is stored.

When the word defining means 21 receive identification information delivered to them, they are arranged for erasing this word sequence information WFI and its transition probability information UWI on the first memory location featured by the identification information II, because this word sequence information WFI was stored as word information WI in the lexicon memory 10. The word defining means 21, when receiving identification information II applied thereto, are further arranged for storing the word information WI of the word sequence defined as one word in the word sequence information WFI of the further

memory locations of the speech model memory 11 featured by the identification information II.

As a result, the advantage is obtained that all the three-word word sequences stored in the speech model memory 11, which are word sequences containing two words and having a typical formulation, are stored in the speech model memory 11 as a word sequence containing only two words.

Next an optimization operation of the optimization device 19 will be described. Fig. 4 shows a third table 24 which contains word information WI and phoneme information PI(WI) which is stored in the lexicon memory 10 after the optimization operation of the optimization device 19. Fig. 5 shows a fourth table 25 which contains the word sequence information WFI and transition probability information UWI which is stored in the speech model memory 11 after the optimization operation of the optimization device 19.

When the information of the first table 17 is stored in the lexicon memory 10 and the information of the second table 18 in the speech model memory 11, and the calculation means 9 deliver activation information AI to the test means 20, the optimization device 19 starts the optimization operation. The test means 20 then test what transition probability information UWI stored in the speech model memory 11 has a value greater than or equal to the minimum value MW. During this test the test means 20 establish that transition probability information UWI of word sequence information WFI = A+C stored on the third row of the second table has the value 10%. The word sequence "international business" therefore represents a typical formulation. After this, the test means 20 deliver the word sequence information WFI = A+C both to the word defining means 21 and to the determining means 22.

After receiving the word sequence information WFI = A+C, the word defining means 21 store the word sequence "international business" in the lexicon memory 10 as one word by means of word information WI = G indicated on the last row of the third table 24. A combination of the phoneme information PI(A) and PI(C) of the words "international" and "business" is assigned to the word information WI = G as phoneme information PI(G) and stored in the lexicon memory 10.

The determining means 22 search for the word sequence information WFI = A+C received from the test means 20 in the word sequence information WFI stored in the speech model memory 11, which is contained in the second table 18. The word sequence information WFI = A+C is found in the stored word sequence information WFI = A+C, WFI = A+C+D and WFI = E+A+C by the determining means 22. The determining means 22 then deliver respective identification information II to the word defining means 21.

Upon receipt of this identification information II, the word defining means 21 are arranged for erasing the word sequence information $WFI = A+C$ still appearing on the third row of the second table 18 and already erased in the fourth table 25 defined as one word having the word information $WI = G$, because this word sequence defined as one word no longer forms a word sequence. When the identification information II is received, the word defining means 21 are further arranged for replacing the word sequence information $WFI = A+C$ by the word information $WI = G$ in the word sequence information $WFI = A+C+D$ and $WFI = E+A+C$ contained in the second table 18 in the sixth and seventh rows, so as to obtain word sequence information $WFI = G+D$ and $WFI = E+G$ represented in the fourth table 25 on the fifth and sixth rows.

As a result of the optimization operation described above, the advantage is obtained that the trigrams $WFI = A+C+D$ and $WFI = E+A+C$ were stored as bigrams $WFI = G+D$ and $WFI = E+G$ in the speech model memory 11 and that a speech recognition operation of these word sequences is possible with less calculation circuitry of the calculation means 9. Furthermore, the advantage is obtained that transition probability information UWI of the word sequence information $WFI = A+C+D$ and $WFI = E+A+C$ contained in the second table 18 continues to be assigned to the word sequence information $WFI = G+D$ and $WFI = E+G$, even after the vocabulary has been optimized, and that no information already determined is lost as a result of this.

The speech recognition device 2 has training means 23 to which word information WI of a recognized text can be applied after a speech recognition operation, and which training means 23 are arranged for extending a word sequence stored in the speech model memory 11 by a word often occurring in a recognized text before or after this word sequence, and for storing transition probability information UWI of the extended word sequence in the speech model memory 11, if the number of words that can be stored for each word sequence in the speech model memory 11 so permits. Since bigrams, having two words per word sequence, and trigrams, having three words per word sequence, can be stored in the speech model memory 11, the training means 23 are arranged for extending a bigram by a word often occurring in a recognized text before or after a bigram.

When the training means 23 detect, for example, that before the word sequence "international business connection" stored as a bigram in the speech model memory 11 as the word sequence information $WFI = G+D$, the word "the" having the word information $WI=E$ relatively often occurs in texts recognized by the calculation means 9, the training means 23 are arranged for storing a word sequence "the international business connection" indicated on

the seventh row of the fourth table 25 and having the word sequence information $WI = E+G+D$ and associated transition probability information UWI of 3% determined by the training means 23.

As a result, the advantage is obtained that a word sequence reduced from a
5 trigram to a bigram in the speech model memory 11 by the word defining means 21 during an optimization operation, can be extended by the training means 23. Consequently, word sequences of typical formulations having a high probability of occurrence can be stored in the speech model memory 11, which word sequences can contain considerably more words than can be stored in a word sequence of the speech model memory 11. As a result, the range of the
10 speech model is increased and typical formulations can be recognized considerably better during a speech recognition operation. As only the word sequences are selected that have a high probability of occurrence and as only these word sequences stored as words are extended beyond the maximum number of three words per word sequence defined for the speech model memory 11, there is advantageously achieved that the required memory capacity in the speech
15 model memory 11 is considerably smaller than when all the possible combinations of, for example, a maximum of four words per word sequence of words stored in the lexicon memory 10 would be stored in the speech model memory 11.

When the training means 23 detect, for example, that in recognized texts the word sequence "International Business Machines Corporation", of which the words are written
20 with initial capitals because they indicate a company name, occur relatively often, the four words can be stored as three word information signals $WI = H$ ("International Business"), $WI = I$ ("Machines") $WI = J$ ("Corporation") in the lexicon memory 10. This word sequence can then be stored as a trigram in the speech model memory 11 under word sequence information $WFI = H+I+J$.

25 This provides the advantage that although only word sequences with a maximum of three words can be stored in the speech model memory 11, the word sequence "International Business Machines Corporation" having four words can very well be recognized during a speech recognition operation and, in addition, also the initial capitals of the words of the word sequence can be detected.

30 Fig. 6 shows a personal computer 26 in which a vocabulary generator 27 in accordance with a second example of embodiment of the invention is realized. The vocabulary generator 27 is arranged for generating and storing word information WI that forms the vocabulary of a speech recognition device. The vocabulary generator 27 has an input terminal

28 which forms input means and at which word information WI of stored text information can be applied as electric input signals to the vocabulary generator 27.

The personal computer 26 has a hard disk 29 which is connected to the input terminal 28. The hard disk 29 contains much text information and many documents
5 respectively, generated, for example, with a text processing program. These documents are formed, for example, by letters, messages or other publications. The contents of these documents relate to a certain domain for which the vocabulary generator 27 is to generate a vocabulary. A certain domain may be, for example, the domain of radiology, botanics or nuclear physics.

10 The input terminal 28 of the vocabulary generator 27 is further connected to the Internet and through the Internet to the memory means 30 which may be formed, for example, by a data server of a university. The personal computer 26 is arranged, in a manner not further shown in Fig. 6, for delivering to an Internet search machine search words which state the specific domain. To the input terminal 28 are connected memory means 30 of the Internet in
15 which documents of the specific domain are stored.

The vocabulary generator 27 includes a storage stage 31 in which the word information WI can be stored. The vocabulary generator 27 is arranged for reading word information WI of text information stored on the hard disk 29 and in the memory means 30 of the Internet and for storing this word information WI in the storage stage 31. The storage stage
20 31 contains many documents or much text information respectively, of the specific domain.

The vocabulary generator 27 includes generating means 32 which are arranged for generating a vocabulary while at least word information WI of the text information of a first and a second word and transition probability information UWI, which indicates the probability of occurrence of the second word after the first word in a word sequence of the text
25 information and is stored in the storage stage 31, can be defined, and which are arranged for storing at least the first and the second word as word information WI in a lexicon memory 10 and for storing the transition probability information UWI in a speech model memory 11. For this purpose, the generating means 32 test all the relevant documents stored in the memory stage 31 about what words very often occur in the specific domain and store these words as
30 word information WI in the lexicon memory 10. These words form the vocabulary of the specific domain.

The vocabulary generator 27 further includes a background lexicon memory 33 in which much word information WI and assigned phoneme information PI(WI) of a general vocabulary is stored. When the generating means 32 detect a word that often occurs in the

documents stored in the storage stage 31 and have stored it in the lexicon memory 10, the generating means 32 search for this word in the background lexicon memory 33. When this word is found in the background lexicon memory 33, the generating means 32 determine the assigned phoneme information PI(WI) of this word on the basis of the background lexicon
5 memory 33, and store the phoneme information PI(WI) of this word assigned to its word information WI in the lexicon memory 10. When this word cannot be found in the background lexicon memory 33, the generating means 32 calculate phoneme information PI(WI) for this word in accordance with statistical methods and store the calculated phoneme information PI(WI) of this word assigned to its word information WI in the lexicon memory 10.

10 The generating means 32 further test what word sequences often occur in the documents stored in the storage stage 31. The generating means 32 store word sequences often occurring as bigrams and trigrams in the speech model memory 11. Transition probability information UWI of the bigrams and trigrams determined by the generating means 32 assigned to the bigrams and trigrams is stored in the speech model memory 11.

15 When the vocabulary generator 27 has finished generating a vocabulary for a certain domain, word information WI and phoneme information PI(WI) is stored in the lexicon memory 10, for example, in accordance with the information contained in the first table 17 and word sequence information WFI and transition probability information UWI is stored in the speech model memory 11, for example, in accordance with the information contained in the
20 second table 18.

The vocabulary generator 27 has an optimization device 19 which is used for optimizing a vocabulary generated by the vocabulary generator 27. The optimization device 19 of the second example of embodiment of the invention here corresponds to the whole optimization device 19 of the first example of embodiment of the invention.

25 When the vocabulary generator 27 has finished generating a vocabulary, the generating means 32 apply activation information AI to the test means 20 of the optimization device 19, after which the optimization operation described with reference to the first example of embodiment commences and the vocabulary is optimized.

30 As the optimization device 19 is included in the vocabulary generator 27, the advantage is obtained that a vocabulary generated by the vocabulary generator 27 is optimized and typical formulations can be better recognized in the documents stored in the storage stage 31 when a speech recognizer that uses the vocabulary generated by the vocabulary generator 27 carries out a recognition operation. Furthermore, in an advantageous manner, the calculation circuitry for recognizing a typical formulation in speech information SI of a spoken

text is considerably smaller with a speech recognition operation of a speech recognizer that uses the vocabulary generated by the vocabulary generator 27.

The generating means 32, after at least a third word has been defined and stored by the word defining means 21, is arranged for extending a word sequence stored in the speech model memory 11 by one word and including the third word, which one word often occurs before or after this word sequence in stored text information applied as an input signal to the generating means 32, and for storing the extended word sequence in the speech model memory 11 if the number of words that can be stored per word sequence in the speech model memory 11 so permits. After an optimization operation of the optimization means 19, text information or documents respectively, stored in the storage stage 31 are again tested by the generating means 32 and a word often occurring in these documents before or after a bigram stored in the speech model memory 11 is stored in the speech model memory 11 as a trigram, which is a combination of this word and the bigram. Transition probability information UWI of the trigram is determined by the generating means 32 and stored in the speech model memory 11 while being assigned to the trigram.

This achieves the advantage that the so-called range of the speech model is further extended and typical formulations are even better recognizable with a speech recognizer that uses the vocabulary generated and optimized by the vocabulary generator 27.

It may be observed that the test means 20 of the vocabulary generator 27, after a positive result of the test of the transition probability information UWI of a word sequence, are also used for determining occurrence probability information, how often this word sequence occurs in the stored text information, and for further testing whether the determined occurrence probability information has a minimum-occurrence value and could not be arranged for defining the words of this word sequence as a third word until there is a positive result of the further test. Such an optimization device with such test means would then advantageously contain only those word sequences, whose transition probability information UWI has a minimum value, as one word each, which word sequences also relatively often occur in stored text information and thus also relatively often in speech information SI of a spoken text. A minimum-occurrence value may be, for example, a value of 2%, which indicates that in 100 words of stored text information the words of the word sequence occur twice as a word sequence.

It may be observed that, for example, the typical formulation "Sehr geehrte Damen und Herren" can be stored in a speech model memory by word defining means during

an optimization operation, and can be recognized very well with little calculation circuitry during a next speech recognition operation.

It may be observed that the optimization device according to the invention can be used very well for optimizing a vocabulary containing composite components, by which a
5 range of the speech model extending beyond a composite built from composite components is achieved.

CLAIMS

1. An optimization device (19) for optimizing a vocabulary of a speech recognition device (2), comprising a lexicon memory (10) in which the word information (WI) of at least a first and a second word forming the vocabulary of a speech recognition device (2) can be stored, and comprising a speech model memory (11) in which at least a probability of occurrence of the second word after the first word in a word sequence formed by these words can be stored as transition probability information (UWI), and comprising word defining means (21) for defining a third word and for storing in the lexicon memory (10) the third word as word information (WI) and for storing in the speech model memory (11) at least transition probability information (UWI) of the probability of occurrence of the third word in a word sequence after at least the first or the second word stored in the lexicon memory (10), characterized in that test means (20) are provided which are arranged for testing whether transition probability information (UWI) of a word sequence stored in the speech model memory (11) has a minimum value (MW), and in that the word defining means (21) are arranged for defining the words of this word sequence as a third word when the test means (20) give a positive test result.

2. An optimization device (19) as claimed in claim 1, characterized in that determining means (22) are provided which, in the event of a positive result of the test of the test means (20), are arranged for determining, as appropriate, at least transition probability information (UWI) already stored in the speech model memory (11), which transition probability information indicates the probability of occurrence of a specific word stored as word information (WI) in the lexicon memory (10) before or after respectively, the word sequence defined as the third word, which determined transition probability information (UWI) can be stored in the speech model memory (11) as a probability of occurrence of specific words before or after respectively, the third word in a word sequence.

3. A speech recognition device (2) for recognizing phoneme information (PI) contained in speech information (SI) of a spoken text and for delivering word information (WI) of a recognized text, comprising input means (3), to which speech information (SI) of a

spoken text can be applied as input signals, and comprising speech recognition means (6) for recognizing phoneme information (PI) of the spoken text contained in the input signals and for delivering word information (WI) of a recognized text, and comprising output means (15) from which word information (WI) of a recognized text can be delivered as an output signal, characterized in that an optimization device (19) in accordance with claim 1 is provided.

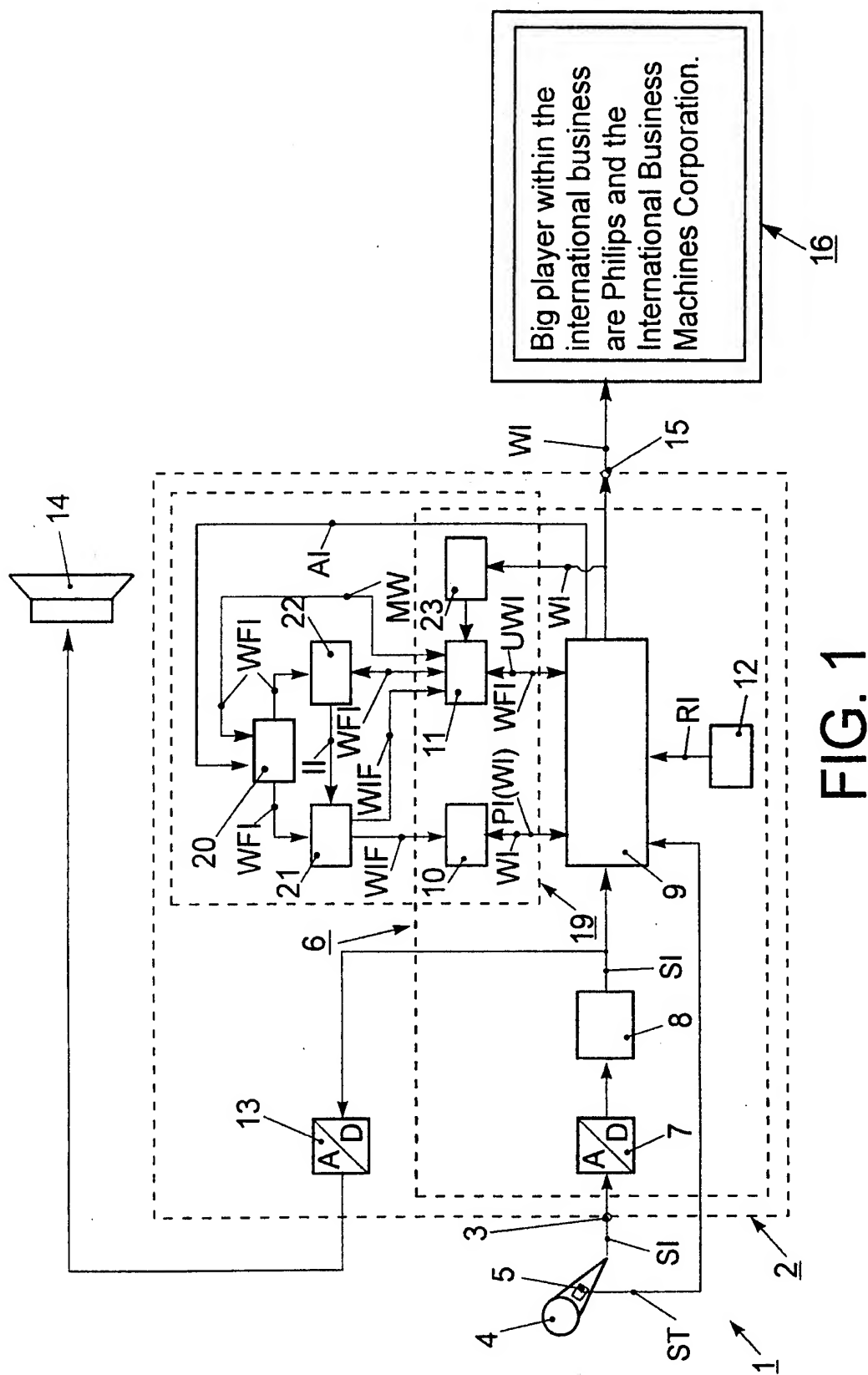
4. A speech recognition device (2) as claimed in claim 3, characterized in that training means (23) are provided to which can be applied word information (WI) of a recognized text and which are arranged for extending a word sequence stored in the speech model memory (11) by a word before or after a word often present in recognized text together with this word sequence and for storing transition probability information (UWI) of the extended word sequence in the speech model memory (11), if the number of words per word sequence that can be stored in the speech model memory (11) so permits.

5. A vocabulary generator (27) for generating and storing word information (WI) forming the vocabulary of a speech recognition device (2), comprising input means (28) to which word information (WI) of stored text information can be applied as an input signal, and comprising generating means (32) which are arranged for generating a vocabulary, at least word information (WI) of the text information of a first and a second word, and transition probability information (UWI) indicating the probability of occurrence of the second word after the first word in a word sequence of the text information, and which are arranged for storing at least the first and the second word as word information (WI) in a lexicon memory (10) and the transition probability information (UWI) in a speech model memory (11), characterized in that an optimization device (19) as claimed in claim 1 is provided.

6. A vocabulary generator (27) as claimed in claim 5, characterized in that the generating means (32), after at least a third word has been defined and stored by the word defining means (21), are arranged for extending a word sequence including the third word, which word sequence is stored in the speech model memory (11), by a word that often occurs before or after this word sequence in stored text information applied as an input signal to the generating means (32), and arranged for storing the extended word sequence in the speech model memory (11) if the number of words per word sequence that can be stored in the speech model memory (11) so permits.

7. A vocabulary generator (27) as claimed in claim 5, characterized in that, after the positive result of the test of the transition probability information (UWI) of a word sequence, the test means 20 are arranged for determining probability-of-occurrence
- 5 information, how often this word sequence occurs in the stored text information and for further testing whether the determined probability-of-occurrence information has a minimum-occurrence value and are not arranged for defining the words of this word sequence as a third word until the result of the further test is positive.

1/3



2/3

WI	PI(WI)
A („international“)	PI(A)
B („machines“)	PI(B)
C („business“)	PI(C)
D („connection“)	PI(D)
E („the“)	PI(E)
F („corporation“)	PI(F)

← 17

FIG. 2

WFI	UWI
A + B	3%
A + C	10%
C + D	4%
C + F	1%
A + C + D	2%
E + A + C	5%

← 18

FIG. 3

WI	PI(WI)
A („international“)	PI(A)
B („machines“)	PI(B)
C („business“)	PI(C)
D („connection“)	PI(D)
E („the“)	PI(E)
F („corporation“)	PI(F)
G („international business“)	PI(G)

← 24

FIG. 4

3/3

WFI	UWI
A + B	3%
C + D	4%
C + F	1%
G + D	2%
E + G	5%
E + G + D	3%
G + B + F	4%

← 25

FIG. 5

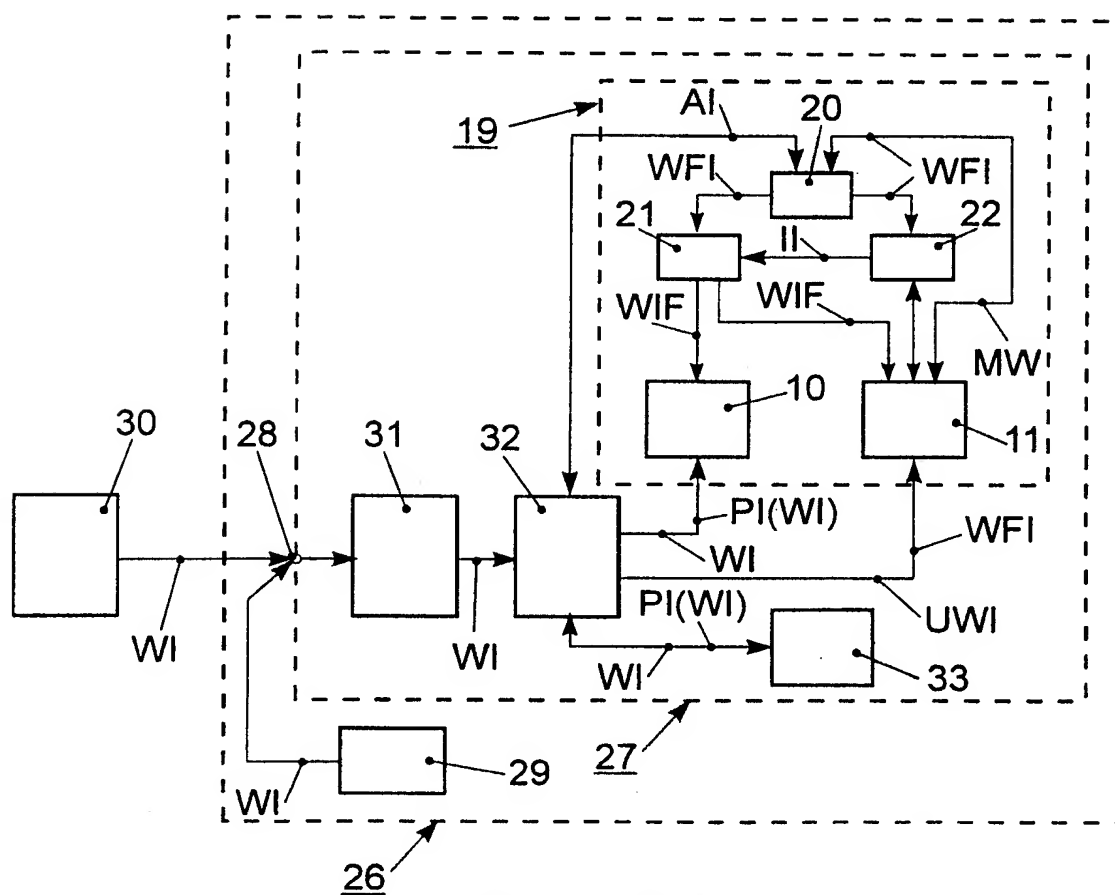


FIG. 6

INTERNATIONAL SEARCH REPORT

Int'l Application No

PCT/EP 99/08640

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 G10L15/18

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	RIES K ET AL: "Class phrase models for language modeling" PROCEEDINGS ICSLP 96. FOURTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING (CAT. NO.96TH8206), PROCEEDING OF FOURTH INTERNATIONAL CONFERENCE ON SPOKEN LANGUAGE PROCESSING. ICSLP '96, PHILADELPHIA, PA, USA, 3-6 OCT. 1996, pages 398-401 vol.1, XP002130925 1996, New York, NY, USA, IEEE, USA ISBN: 0-7803-3555-4 paragraph '0002! — —/—	1-7

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.*** Special categories of cited documents:****"A"** document defining the general state of the art which is not considered to be of particular relevance**"E"** earlier document but published on or after the international filing date**"L"** document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)**"O"** document referring to an oral disclosure, use, exhibition or other means**"P"** document published prior to the international filing date but later than the priority date claimed**"T"** later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention**"X"** document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone**"Y"** document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.**"Z"** document member of the same patent family

Date of the actual completion of the international search

18 February 2000

Date of mailing of the international search report

06/03/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Krembel, L

INTERNATIONAL SEARCH REPORT

Int'l Application No
PCT/EP 99/08640

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KLAKOW D: "Language-model optimization by mapping of corpora" PROCEEDINGS OF THE 1998 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, ICASSP '98 (CAT. NO.98CH36181), PROCEEDINGS OF THE 1998 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, SEATTLE, WA, USA, 12-1, pages 701-704 vol.2, XP002124207 1998, New York, NY, USA, IEEE, USA ISBN: 0-7803-4428-6 paragraph '0003!	1-7
X	GIACHIN E P: "PHRASE BIGRAMS FOR CONTINUOUS SPEECH RECOGNITION" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP),US,NEW YORK, IEEE,1995, pages 225-228, XP000657971 ISBN: 0-7803-2432-3 paragraphs '0003!, '0004!	1,3,5
P,X	EP 0 935 238 A (PHILIPS PATENTVERWALTUNG ;KONINKL PHILIPS ELECTRONICS NV (NL)) 11 August 1999 (1999-08-11) page 2, line 40 - line 54	1-7
A	HWANG K: "VOCABULARY OPTIMIZATION BASED ON PERPLEXITY" IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP),US,LOS ALAMITOS, IEEE COMP. SOC. PRESS,1997, pages 1419-1422, XP000822723 ISBN: 0-8186-7920-4 paragraph '0002!	1,3,5
A	DELIGNE S ET AL: "LANGUAGE MODELING BY VARIABLE LENGTH SEQUENCES: THEORETICAL FORMULATION AND EVALUATION OF MULTIGRAMS" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING (ICASSP),US,NEW YORK, IEEE,1995, pages 169-172, XP000657957 ISBN: 0-7803-2432-3 paragraph '0002!	1,3,5

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 99/08640

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0935238 A	11-08-1999	DE 19804603 A	12-08-1999
		JP 11272290 A	08-10-1999